DOCUMENT RESUME

ED 084 434 CE 000 648

AUTHOR McKinlay, Bruce
TITLE Validity and Readability of the Occupational Information Access System "QUEST" Questionnaire.
PUB DATE [71]
NOTE 55p.; For related documents, see CE 000 649 and CE 000 652

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Computer Oriented Programs; *Information Systems; *Information Utilization; Occupational Information; Questionnaires; *Readability; Reading Level; Use Studies; *Validity
IDENTIFIERS *Occupational Information Access System

ABSTRACT

The purpose of the report was to examine the readability of the Occupational Information Access System (OIAS) QUEST questionnaire and the validity of independent client use of the questionnaire as an information gathering device. Readability formulas rate the QUEST questionnaire "easy" or "fairly easy" and they rank the occupational descriptors "difficult" or "very difficult," as they do other occupational information materials. In field testing, however, users rated the system as easy or very easy to use with wording problems reported on only two percent of the QUEST questions answered; only one-half of the percent of the questions were wording problems resulting in response errors. A readability and validity test in three Employment Service offices indicated that client self-reports of preferences and abilities concurred with counselor estimates about 80 percent of the time. There is some indication in this research that client self-reports are as valid as counselor estimates, and other research summarized in this report indicates that self-reporting is as valid as testing. It was concluded that although self-reporting may not be relied upon completely for all clients, it is at least one of the appropriate bases for questionnaire response. (Author/SC)

ED 084434

VALIDITY AND READABILITY OF THE OCCUPATIONAL INFORMATION ACCESS SYSTEM "QUEST" QUESTIONNAIRE

by

Bruce McKinlay

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Purpose of the Report

The purpose of the report is to examine the readability of the Occupational Information Access System QUEST questionnaire and the validity of independent client use of the questionnaire as an information gathering device. These are legitimate issues in the evaluation of any such program, and they take on added significance when the system is considered for use with disadvantaged persons.

Abstract

"Readability formulas" rate the QUEST questionnaire "easy" or "fairly easy" (7th to 10th grade reading level) to read, and they rank the occupational descriptions "difficult" or "very difficult", as they do other occupational information materials. But readability formulas are very poor devices for measuring the comprehensibility of written material for disadvantaged adults.

Field testing of these materials showed very few readability problems. Over 90 percent of the counselors and clients in various schools and social

CE000 648

FILMED FROM BEST AVAILABLE COPY

agencies who tested the System rated it easy or very easy to use. In a special readability and validity test conducted in three Employment Service offices, wording problems were reported on only two percent of the QUEST questions answered, and in only one half of one percent of the questions were wording problems associated with response errors. Surprisingly disadvantaged users reported fewer wording problems than non-disadvantaged users.

Even with the more difficult occupational descriptions 96 percent of the users (in this case high school students) said they were easy or very easy to read.

Readability does not seem to be much of a problem with the Occupational Information Access System. The crucial issue has to do with the validity of self-reporting. When and by whom may the questionnaire be used independently, and when should it be used as a tool by a counselor? The special readability and validity test in three Employment Service offices indicate that clients self-reports of preferences and abilities concur with counselor estimates about 80 percent of the time, with disadvantaged clients having no more trouble than non-disadvantaged clients. (People of low ability do tend to have more trouble self-reporting, however.)

There is some indication in this research that clients self-reports are as valid as counselor estimates, and other research, summarized in this report, indicates that self-reporting is as valid as testing. Thus, while self-reporting may not be relied upon completely for all clients, it is at

least one of the appropriate bases for questionnaire response. Additionally, client self-reporting provides a reflection of his subjective self-perception, and that in itself is a helpful tool in counseling.

Scope of the Research

The following tests have been performed, and the results are reported below. (1) Readability formulas have been applied to several components of the OIAS system, (2) information regarding "ease of use" ratings by both clients and counselors have been extracted from the first system evaluation, (3) the questionnaire has been used in actual counseling interviews in three Employment Division offices in Portland Oregon where counselors reviewed the questionnaire with their clients specifically to identify ambiguous or difficult items, (4) client response to questionnaire items has been validated against counselor assessments and against GATB scores, (5) the questionnaire and instructions have been reviewed by two psychologists experienced in construction of vocational assessment instruments.

The OIAS system is designed for independent use by clients and students. This feature does not preclude client use directed by a counselor or use exclusively by a counselor. (In fact, features have been built into the system to facilitate these types of uses.)

Three things are essential if the client is to use QUEST himself. (1) He must be able to understand the question. (2) He must be able to answer the question within some range of acceptable error. (3) He must be willing to answer the question, i.e. he must have incentive to report his abilities and

preferences accurately. It is an advantage if the user also sees the relevance of the questions to his career choice. (That is one of the goals of the Occupational Information Access System QUEST program which the system seems to meet.)

The first of these factors is a question of comprehensibility and will be discussed in Part I of this report. The questions of ability and willingness to report are questions of validity in self-reporting which will be examined in Part II.

Part I: Readability

The usefulness of OIAS or any other labor market information intended for client use depends in part upon its comprehensibility to clients. It should pass a comprehensibility test.

Unfortunately, there are no "comprehensibility tests" as such. We are obliged, therefore, to base our judgments upon some rough measures of readability (readability formulas) and upon experience with the material.

Readability Formulas

Readability formulas are widely used by publishers of elementary school text books and by reading teachers to determine the reading level (usually expressed in grade-level terms) of printed materials. There are more than 30 different readability formulas, of which about a half dozen are in common usage. Several of these formulas were reviewed for applicability to the questionnaire at hand, and two were selected: the widely used Dale-Chall formula and the Flesch formula, which avoids a "word list" as a way of

determining word difficulty.

Fundamentally, readability formulas attempt to measure difficulty of reading style. In doing so, they take into account two factors: word difficulty and sentence difficulty. Words are classified as easy in the widely used Dale-Chall formula if they appear on a list of 3,000 words commonly read by sixth grade students; they are classified as hard if they do not appear on that list. In the Flesch formula, a word's difficulty is measured by the number of syllables it contains. In most formulas, including both Dale-Chall and Flesch, sentence difficulty is measured by the average number of words per sentence.[1] Thus, according to readability formulas, the shorter the word and the shorter the sentence, the easier the reading.

Some notation of the applicability of readability formulas is in order, both because of their inherent limitations and because of the unusual way in which we are using them.

---

[1] In the Dale-Chall formula:
Raw Score = 3.6365 + .1579 D.S. + .0496 a.s.l.
where, D.S. (Dale Score)= percent of words not on the Dale list, and
a.s.l. (average sentence length)= average number of words per sentence.
The Flesch formula is of the form:
R.E. (reading ease)= 206.835 - .846 w.l. - 1.015 s.l.
where, w.l. (word length)= average number of syllables per 100 words, and
s.l. (sentence length)= average number of words per sentence.
The resulting scores -- calculated to three and four decimal places -- are then translated to grade levels or narrative descriptors of difficulty.

The formulas are limited measures of readability.

> They measure readability only in a general way, for the general reader. They do not measure many of the factors that go into good writing..... Even those factors that readability formulas purport to measure are not measured in an absolute way. The shortest word is not always the simplest; the most "familiar" word is not always the best; the shortest sentence is not always the clearest or the most readable.[2]

> The oft-repeated criticism that formulas lead to "primerized" writing stems from this implication, since perhaps the most obvious characteristics of primers is their short, simple sentences. Primers must have short sentences, probably nearly all short sentences, but books for adults, even those on unfamiliar topics, are not primers in this sense. The adult brings a background to his reading that helps him understand longer sentences.[3]

Excessive reliance on readability can lead to material that is not so much easy to read as it is childish.

Readability research has demonstrated that content, stylistic elements, format, and organization all contribute to difficulty. So far, only stylistic elements have been amenable to reliable quantitative measurement and verification.

> The problem of validity is difficult. First of all, there are no rigorous standards of just what is 4th grade difficulty as opposed to 5th grade difficulty. There seems to be some loose sort of agreement between publishers and educators which is based on experience and perhaps a little on test data as to what grade level designations mean. However, even standardized

---

[2]George R. Klare and Byron Buck, Know Your Reader, (New York: Hermitage House, 1954), p. 137.

[3]Ibid., p. 139

> test data are not exact.... In general, newer tests are more difficult or, in other words, a 9th grade student today reads better than a 9th grade student in former years.
>
> .....
>
> Hence the problem of validity is complicated by trying to determine grade level when grade level won't stand still and when subjective "judgments" are about as good a standard as can be found.....[4]

The formulas are probably more valid when used ordinally to assess the comparative difficulty of various pieces of written material than they are when used in an attempt to establish an absolute reading level.

A further limitation for our purposes is the fact that readability formulas are geared to elementary school education. One readability formula developer says frankly of his formula "Grade level designations are determined by simply plotting lots of books which publishers said were 3rd grade readers, 5th grade readers, etc."[5] What a readability formula really says is that the material in question is written like text books said to be for children in a particular grade.

Even if some consistency in the measurement of readability could be achieved, improved readability does not assure increased comprehension, which is the real issue here. Some research has been done on this matter. "The general conclusion to be drawn from this work is that the writer

---

[4]Edward Fry, "A Readability Formula That Saves Time," Journal of Reading, XI, No. 7, (April, 1968), p. 515.

[5]Ibid., p. 515.

certainly cannot always count on improvements due simply to the use of more readable materials."[6]

What a fourth-grade child can read is not necessarily the same thing a disadvantaged adult can read. "Disadvantaged" is an economic and social designation, not an educational or intelligence measure, and clients of manpower programs are youths and adults, not children. What is appropriate for one is not necessarily appropriate for the other.

There are several important factors besides readability which account for a reader's ability to comprehend written material, namely: the reader's educational level, the amount of time the reader has to spend, the reader's background and experience with a given topic, and the kind and level of the reader's motivation. These factors are probably especially significant for adults who are poor readers, but there has been practically no research into the applicability of readability formulas to poor reading adults. The only study of readability for adults of low reading ability which we were able to locate was published in 1934.[7] Entitled "A Study of the Factors Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability," this 1934 study correlated 29 factors of reading materials with the subject

---

[6]George R. Klare, The Measurement of Readability, (Ames, Iowa: Iowa State University Press, 1963), p. 16.

[7]Inquiries to two Educational Resources Information Center (ERIC) Clearinghouses revealed no recent research on the subject. That fact is remarkable in itself, considering the current concern with the education of disadvantaged adults.

adult readers' comprehension score on sample materials (Medical and health articles). "The factor most closely related to difficulty of comprehension was the number of different words in the selection. The factor next most closely related to difficulty was the number of hard, non-technical words in the selection...."[8] The coefficients of correlation between the factor and comprehension scores for these items was -.462 for number of different words and -.380 for number of hard, non-technical words. As a rule, direct measures of word difficulty showed higher degrees of significance than factors having to do with word length, sentence length, and sentence structure. The two factors used in most readability formulas, word length and sentence length, ranked fairly low.[9]

The inherent limits of readability formulas together with this empirical evidence raise serious doubt whether readability formulas validly measure adult readers' (to say nothing of disadvantaged adult readers') ability to comprehend written material. Thus the use of readability formulas for our

---

[8]Edgar Dale and Ralph W. Tyler, "A Study of the Factors Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability," Library Quarterly, IV, (July, 1934), p. 399.

[9]The average number of words in each sentence of the selection showed a correlation coefficient of -.293 and ranked 11th among the 29 variables; the percentage of bisyllabic words in the selection showed a -.227 correlation coefficient and ranked 13th; while the average difficulty score for the technical words in the selection as determined by the Dale list showed a .225 correlation coefficient and ranked 16th.
Source: Dale and Tyler "Factors Influencing Difficulty for Adults" Table 1, p. 400.

purpose is suspect, and formula results must be viewed as only a rough measure at best.

Table 1 reports the scores resulting from application of the Flesch and Dale-Chall formulas to several OIAS items. Of the two, the Flesch formula appears to be the more appropriate for the purposes at hand, because it measures word difficulty on the basis of number of syllables rather than by a prescribed word list which favors elementary school text book vocabulary and contains very few job related words. For instance, the Dale list rates "playground," "neighborhood," and "granddaughter" as easy

TABLE 1

READABILITY OF OIAS QUESTIONNAIRE

ITEM	FLESCH "READING EASE" SCORE (STYLE)	DALE-CHALL RAW SCORE (GRADE LEVEL)
Possible Scale Range	100 (easy for any literate person) to 0 (practically unreadable)	3.6365 (4th grade & below to 10.0 & above) (16th-college grad)
Introduction to OIAS Questionnaire	73.778 (Fairly easy)	6.7892 (7-8th)
OIAS Questionnaire: Question 1-4	85.030 (Easy)	7.3785 (9-10th)
OIAS Questionnaire: Questions 1, 4, 8 9, 14, 17, 20, 25	47.1713 (Difficult)	7.5430 (9-10th)

words, but classified "employer," "firm," "openings," and "experience" as hard words. The elementary school orientation of the Dale-Chall list tends to exaggerate the difficulty of any labor market information relative to elementary school text books because of the subject matter, regardless of the reading difficulty for the population actually using the material.

The Flesch formula rates the introductory paragraphs of the OIAS questionnaire as "fairly easy," while rating the first four questions themselves "easy." A sampling of questions from throughout the questionnaire shows an appreciably lower (more difficult) rating by the Flesch Formula. (See Table 1.) The Dale-Chall formula, on the other hand, says the material is harder over all and contradicts Flesch as to which is the easier to read. The Flesch and Dale-Chall readability formulas thus give somewhat contradictory assessments of the difficulty of the questionnaire.[10]

An interesting adjunct to the Flesch reading ease formula is a method for calculating "human interest" scores based on the proportion of "personal words" (personal pronouns, proper names, etc.) and "personal sentences" (sentences addressed to the reader, etc.). On this scale the questionnaire introduction rated "interesting."

---

[10]Studies comparing results when formulas are applied to novels and text books have regularly shown high degrees of correlation among the results of the various formulas. The disparities here are evidence of the lack of concurrent validity of readability formulas when applied to occupational information.

User Evaluation

Experience with the questionnaire provides other and perhaps more valid evaluations. At the end of the first OIAS test period both counselors and clients were asked to evaluate the system and suggest changes. There was only one comment in any way related to the comprehensibility of the questionnaire or other system materials from the 40 clients included in the evaluation (an admittedly small sample, but one known to include "potential dropouts" from high school, "unmotivated" Vocational Rehabilitation clients, and a variety of Employment Division counseling clients). That comment said that "the questionnaire should be more specific." It appears from the context of the comment that it had reference to questionnaire output i.e., the number of occupational titles, rather than to questionnaire wording.

All of the clients in the evaluation were asked to rate system usage on a four-point scale. Their responses reflect on the readability of the materials along with the clarity of instructions, operational difficulty of mechanical devices, etc. In assessing the system in general, 23 of the 40 clients said it was "very easy" to use, 16 said it was "easy", one said it was "difficult" and none rated it "very difficult."

Counselor reaction to the system was evaluated independently. Of the 22 counselors included in the evaluation, 12 rated the system "very easy" to use, 6 said it was "easy," three said it was "difficult," and one did not answer the question. Among the counselors' various suggestions for change were two which were directed at the questionnaire. These were "revise or redefine the questionnaire," and "eliminate the questionnaire." These comments

were directed at potential problems of self-reporting and of independent client usage, not particularly at the readability of the material.

In sum, the user evaluation identified no problems with the readability of OIAS material. Neither clients nor counselors complained of trouble understanding the system, nor did they make suggestions for clarifying or simplifying it. Over 90 percent of the evaluations received from counselors and clients rated the system easy or very easy to use.

A subsequent evaluation by high school students in the spring of 1971 yielded similar results. When asked whether the questionnaire was easy to understand, 110 students answered as follows:

HIGH SCHOOL STUDENT EVALUATION OF OIAS QUESTIONNAIRE

Easy to Understand	(n = 110)
YES, definitely	47.3%
YES	46.4%
NO	6.3%
Definitely Not	0%

Ninety-four percent rated it easy to understand, and none thought it was very difficult.

"Quest-Test" for Readability

A second, more specific attempt to identify readability problems was undertaken in three Employment Division offices in Portland, Oregon (The Adult Opportunity Center, the Youth Opportunity Center, and the Concentrated Employment Program) in February 1971. In that "Quest-Test" counselors

were asked to give the questionnaire to clients and to let them complete it independently. The counselors were then instructed: "After the client has completed his questionnaire, please go over it with him to identify any problems he has encountered. Make note of any problems either on the questionnaire itself or on a separate sheet of paper, being as explicit as you can. If possible suggest changes which would correct the problem."[11]

Seventeen of the 61 users in the "Quest-Test" identified wording problems. Those problems appeared on 2 percent of the questions answered. However in only half of one percent of the questions were wording problems associated with response errors, (where response error is defined as a discrepancy between the response of the client and the independent response of his counselor). In the other $1\frac{1}{2}$ percent, the wording problems reported by the clients were not sufficiently serious to cause response errors.

Two questions raised the greater number of wording problems. These were:

Question 4: Region Five clients (8 percent of the total) reported wording problems. All questioned the meaning of the response categories, e.g. "Does Immediate Area mean Portland or just part of Portland?"[12]

---

[11] A full explanation of the Quest-Test procedure appear in Part II of this report.

[12] The Eugene questionnaire read "...Immediate Area (Lane County)." The parenthetical definition was crossed off the questionnaire used in the Portland Quest-Test, but no new area definition was inserted, even though it would be on a Portland area questionnaire.

"Does 'Northwest' mean Pacific Northwest or Northwest Portland?" Such ambiguities can readily be eliminated on the Portland questionnaire.

Question 8: Education Five clients indicated problems. One said interest is more important than education, one did not understand the cumulative character of the response categories, one said she would like to get more education than she indicated she planned to get, one said he was not sure, and one wondered whether the responses were intended to include on-the-job training.

No other question drew more than two complaints, except Question 25, Earnings. Here two people said the question was unclear and one had trouble choosing between two of the response categories.

Ten of the questions showed no wording problems. Surprisingly, the interest and aptitude questions, where questions have been raised about clients' ability to self report, produced few complaints about wording.

The incidence of wording problems among various types of Quest-Test clients is reported in the following table.

The expected pattern is for persons with low educational attainment, low intelligence or low verbal aptitude to have the greatest problems. In fact, the pattern appears to be the reverse, e.g. persons with high verbal aptitudes reported more wording problems than did persons with low verbal aptitudes. However none of these differences is statistically significant. An attempt to correlate wording problems with reading test scores proved to no avail, because only one of the clients for whom reading test scores were available

TABLE 2

QUEST-TEST CLIENTS' CHARACTERISTICS
AND INCIDENCE OF WORDING PROBLEMS

CLIENT CHARACTERISTICS	NUMBER OF CLIENTS IN QUEST-TEST	PERCENT OF QUESTIONS ON WHICH CLIENTS IDENTIFIED WORDING PROBLEMS
TOTAL	61	2.0 %
Highest School Grade:		
0-7	1	8.0
8-11	29	0.8
12	22	1.6
Over 12	9	5.8
Age:		
Under 22	24	1.5
22-54	35	2.2
55+	2	4.0
Disadvantaged:		
YES	42	1.1
NO	19	3.8
Veteran:		
YES	8	2.5
NO	50	2.0
Handicapped:		
YES	8	1.5
NO	50	2.2
Aptitude Test Scores (GATB)		
Intelligence (G)		
Level 1	2	0
Level 2	11	6.5
Level 3	12	1.0
Level 4	11	1.2
Level 5	4	2.0
Verbal (V)		
Level 1	3	1.3
Level 2	8	8.5
Level 3	14	1.1
Level 4	10	1.2
Level 5	5	0.8

also reported a wording problem. The one client characteristic which was significantly correlated with wording problems was disadvantaged status,[13] but the pattern was the reverse of that expected. Disadvantaged clients reported significantly fewer wording problems than did non-disadvantaged clients, and their wording problems were about equally likely to be associated with response errors.

Some explanation of this surprising result is in order. Some of the counselors who administered the Quest-Test felt that their disadvantaged clients were less likely to report problems because they would see such a report as a sign of personal inadequacy. It also appeared that some higher educated, non-disadvantaged clients had trouble because of their high verbal sophistication, while others were over-zealous in citing potential problems rather than problems they themselves actually encountered.

Overall, this evaluation of the readability of the questionnaire is consistent with the other assessments described in this section of the report. It revealed a low number of wording problems (2 percent), many of which were more in the nature of suggestions than actual problems. While they did not identify serious problems, comments made by clients as well as the several suggestions from counselors will be used in adapting the questionnaire for the Portland test. Two meetings were held with counselors who participated in the Quest-Test to solicit their suggestions for improving

---

[13] $x^2 = 4.39$; df = 1; significant at the 0.05 level

wording with the result that the questionnaire used in the Portland test will be clearer in many places than the one used in the Quest-Test.

Readability of OIAS Occupational Descriptions

Though no questions have been raised by anyone regarding the readability of the OIAS descriptions (the 250-word occupational briefs), these are potentially subject to much greater readability problems. Their difficulty comes about because a more sustained reading effort is required, the subject matter is more technical and more varied (hiring requirements, wage rates, job duties, types of training programs, etc.), and the project staff had much less opportunity to do the kind of careful writing and editing which would be desirable for the descriptions, since resources for information development (as opposed to system development) were extremely limited. Predictably, a higher reading level is apparent in those materials.

Table 3 reports the readability scores of two OIAS descriptions along with other, standard occupational information. All ten of the samples rate "difficult" or "very difficult" according to Flesch and, with one exception, require at least 11 or 12th grade reading ability according to Dale-Chall. This finding is consistent with the conclusions of a study done in 1950. That study analyzed 78 pieces of occupational information from various governmental and commercial sources according to the Flesch formula; it found two-thirds ranking "very difficult" and one-third "difficult."[14] The limited

---

[14]Arthur H. Brayfield and Patricia Aepli Reed "How Readable Are Occupational Information Booklets," Journal of Applied Psychology XXXIV, No. 5 (1950), pp. 325-328.

TABLE 3

READABILITY OF OIAS DESCRIPTIONS
AND OTHER OCCUPATIONAL INFORMATION

ITEM	FLESCH "READING EASE" SCORE (STYLE)	DALE-CHALL RAW SCORE (GRADE LEVEL)
J.G.Y.W.: Automobile-Service Station Attendant (Employment Prospects section)	44.593 (difficult)	6.8537 (7-8th)
D.O.T.: Truck Drivers, Light	44.406 (difficult)	8.5901 (11-12th)
J.G.Y.W.: Medical Laboratory Assistant (Qualifications section)	39.839 (difficult)	8.6025 (11-12th)
OIAS Occupational Desc.: Truck Driver (full narrative style)	38.991 (difficult)	8.7374 (11-12th)
OIAS Occupational Desc.: Counselors (full narrative style)	33.265 (difficult)	9.3972 (13-15th)
OIAS Occupational Desc.: Truck Driver (print-out style)	27.318 (very difficult)	9.3262 (13-15th)
O.O.H.: Soil Scientists (Nature of Work section)	22.924 (very difficult)	10.5025 (college grad)
O.O.H.: Hospital Attendants (Employment Outlook section)	17.671 (very difficult)	9.1223 (13-15th)
OIAS Occupational Desc.: Counselor (print-out style)	15.274 (very difficult)	10.2776 (college grad)
D.O.T.: Director of Placement	-3.224 (very difficult)	12.7828 (college grad)

J.G.Y.W. = Job Guide for Young Workers
D.O.T. = Dictionary of Occupational Titles
OIAS = Occupational Information Access System
O.O.H. = Occupational Outlook Handbook

applicability of readability formulas must be borne in mind in assessing these figures, however.

Two versions of the OIAS descriptions are presented in Table 3. The first, labeled "print-out style" refer to the descriptions as they appear over the terminal. The style used in those presentations is characterized by the frequent omission of articles, heavy usage of colons and semi-colons rather than separate sentences, and other stylistic features which compress the material and speed the flow of facts by the omission of non-essential grammatical elements. For instance, the description for counselors reads, in part:

> ...CURRENT LOCAL EMPLOYMENT: 160, 40% MALES AND 20% OVER AGE 54. EMPLOYERS: PUBLIC SCHOOLS, PUBLIC SOCIAL AGENCIES, AND PRIVATE NON-PROFIT ORGANIZATIONS...

Readability calculations were also made for the full, narrative version of the same statements. ("The current local employment is 160, 40 percent of whom are males and 20 percent over age 54. Employers include public schools, public social agencies, and private non-profit organizations...").

The print-out style, though it is preferred by system users, results in a readability score one or two grades higher (according to Dale-Chall) and more difficult (according to Flesch) than the full narrative version. For analysis purposes, both have been included in Table 3.

It is apparent from Table 3 that the OIAS descriptions are somewhat

more difficult than material in the Job Guide for Young Workers, approximately the same level of difficulty as the Occupational Outlook Handbook, but are probably easier reading than the Dictionary of Occupational Titles. It is also apparent that the descriptions are harder to read than the QUEST questionnaire and system instructions.

Nevertheless, there have been no complaints, either in the evaluation or informally, about the readability of the descriptions. On the contrary, many users have been complimentary of the comprehensive, factual style of the descriptions.

Extensive use of the material with a "vocational survey" class of educationally disadvantaged ninth graders presented no readability problems, according to the teacher.[15]

The spring 1971 evaluation at Eugene's Churchill High School produced very favorable comments by students about the readability of OIAS descriptions.

HIGH SCHOOL STUDENT EVALUATION OF JOB DESCRIPTIONS

Easy to Understand	(n = 114)
YES, definitely	35.1%
YES	60.5
NO	3.5
Definitely not	0.9

Fun to Use	(n = 114)
YES, definitely	50.9%
YES	48.2
NO	0
Definitely not	0.9

[15] Students in the class all fell in the 4th and 5th stanines on a composite score of the Iowa Test of Educational Development.

Why does this nominally difficult material apparently present so few readability problems? Several explanations have been advanced by counselors and teachers who have used the system. Motivation is probably one reason. The topic (potential occupational choice) is one of high interest, and such motivation can compensate for difficult material. In this vein, it has been observed that clients and students appreciate the objective, factual, "non-promotional" style of the descriptions.

Length of the statements may be another reason. Several people have suggested that statements of 250 words appear manageable to people when books, even reference books of many similarly short statements, seem formidable.

The fact that the descriptions are usually displayed via teletype terminal (at the rate of 10 characters per second) also appears to be a significant factor, and one which compensates for readability difficulty. Watching the material being printed does stimulate the terminal user to read each word and may therefore tend to increase comprehension.

This area is obviously ripe for research. For instance, which occupational topics are most difficult to communicate to poor readers? Which of the labor market analysts' terms are obscure to job seekers? To what extent do maturity of the reader, motivation to get facts, and shortness of the descriptions, compensate for conceptual and word difficulty? Finally, what effect does the use of the teletype terminal have on the communication?

Unfortunately these and other such interesting questions are beyond the scope of this project. We can only hope to discover, through field testing, whether occupational material of the general content, style, and format used in OIAS is useful to people. Precisely why it is or is not useful will remain unknown without further research. This is indeed unfortunate since improvements could easily be made if the sources of problems were known. Such knowledge would also rationalize the development of materials, whereas occupational information is now prepared in many variations on the assertion that some new style, or format, or medium is going to be effective where others have supposedly not been.

For the present, we must rely on the information we have available. That information indicates that the readability of the Occupational Information Access System questionnaire and related instructions is within the capability of system users. Further, it indicates that even the more difficult occupational descriptions are sufficiently comprehensible to be both useful and attractive to seekers of occupational information.

Part II: Validity of Independent Client Use

There are certain criteria against which a questionnaire should be judged. Essentially they are criteria of clarity, conceptual soundness, and efficacy. In the case of QUEST, certain additional matters of validity must also be considered.

Technical Review

For an opinion on overall questionnaire design the QUEST questionnaire was reviewed by two psychologists with extensive experience and expertise with questionnaire and test design with a particular interest in career choice.[16] After examining the questionnaire, these reviewers judged it to be technically sound. In a subsequent interview they pointed out that with such a questionnaire, one does not have to prove that it predicts or measures anything, only that it identifies some pertinent occupations (a test of efficacy). Further, they observed that validity in the usual sense of the term is perhaps the wrong word to use; rather you should test the questionnaire for utility.

These reviewers further observed that there is much "processing" that goes on in a person's mind when he considers an occupation, and much of it will probably never be measured. The Questionnaire is good if it identifies some new, pertinent occupations. This is a particularly pertinent observation since the questionnaire is intended as a rough-cutting, not a fine-

---

[16]Interview with Leona Tyler, Ph.D., Dean of the Graduate School and Professor of Psychology at the University of Oregon, and Norman D. Sundberg, Ph.D., Dean of the School of Community Service and Public Affairs and Professor of Psychology at the University of Oregon. Dean Tyler has a long history of research in psychology and vocational choice. She is the author of the book The Psychology of Human Differences and numerous articles and reports. Dean Sundberg collaborated with Dean Tyler on a cross-cultural research study of career choice patterns and has professional interests in career choice and methods of professional education.

cutting instrument. The output, as described to the user, is a list of occupations to explore, not the name of the one right occupation for the individual. Both reviewers were favorably impressed with the way the QUEST questionnaire rationalizes the career exploration process by raising significant questions which individuals might otherwise overlook.

Their questions related to three features of the questionnaire. They asked about the basis for the individual questions, especially the interest questions which are based on the data-people-things categories of the D.O.T. This was a matter of special attention because interests as usually defined seem to be a weak factor in career selection. They also wondered, as others have, about self-reporting and about the readability of the materials.

The question of readability has already been dealt with; the remainder of this report is devoted to some further questions of validity in the use of self-reporting as a means of information collection.

Validity

Just what constitutes validity is an issue of long debate and little agreement among psychometricians and users of assessment devices. In 1954 the American Psychological Association, the American Educational Research Association, and the National Council on Measurements Used in Education clarified the issue and introduced some authority to the field by producing and endorsing a set of "Technical Recommendations for

Psychological Tests and Diagnostic Techniques."

These Recommendations identify four kinds of validity measures, the appropriate one(s) depending on the purpose of the test or other diagnostic technique. The four kinds of validity measures recognized in the Recommendations are content, predictive, concurrent, and construct. These are defined as follows:

Content Validity

1. The test user wishes to determine how an individual would perform at present in a given universe of situations of which the test situation constitutes a sample.

a. Content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn. Content validity is especially important in the case of achievement and proficiency measures.

Predictive Validity

2. The test user wishes to predict an individual's future performance (on the test or on some external variable).

b. Predictive validity is evaluated by showing how well predictions made from the test are confirmed by evidence gathered at some subsequent time. The most common means of checking predictive validity is correlating test scores with a subsequent criterion measure. Predictive uses of test include long-range prediction of intelligence measures, prediction of vocational success, and prediction of reaction to therapy.

Concurrent Validity

3. The test user wishes to estimate an individual's present status on some variable external to the test.

c. Concurrent validity if evaluated by showing how well test scores correspond to measures of concurrent criterion performance or status. Studies which determine whether a test discriminates between presently identifiable groups are concerned with concurrent validity. Concurrent validity and predictive validity are quite similar save for the time at which the criterion is obtained. Among the problems for which concurrent validation is used are the validation of psychiatric screening instruments against estimates of adjustment made in a psychiatric interview, differentiation of vocational groups, and classification of patients. It should be noted that a test having concurrent validity may not have predictive validity.

Construct Validity

4. The test user wishes to infer the degree to which the individual possesses some trait or quality (construct) presumed to be reflected in the test performance.

d. Construct validity is evaluated by investigating what psychological qualities a test measures, i.e., by demonstrating that certain explanatory constructs account to some degree for performance on the test. To examine construct validity requires both logical and empirical attack.[17]

A question immediately arises as to whether or not QUEST is a test in the sense used in the Technical Recommendations and therefore subject to measures of validity. Tests are measuring devices designed to determine a person's knowledge, abilities, aptitudes, etc. The questionnaire is not a measuring device in that sense, but an instrument for recording information which is presumed to be known to the individual. Its validity therefore depends upon its readability (analyzed in Part I), upon the users ability and

---

[17] American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education, "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Psychological Bulletin, LI, No. 2, Part 2: Supplement, (March 1954), pp. 1-38.

willingness to answer the questions (analyzed in this part of the report), and upon the validity of the labor market and worker trait factors on which the individual questions are based.

With regard to the types of validity defined in the Technical Recommendations, it is clear that QUEST is not a proficiency test, so content validity need not be demonstrated.

The QUEST program may have predictive value (2, above); it is plausible to argue that persons are most likely to enter occupations where physical demands, working conditions, educational requirements, aptitudes, interests, and compensation are consistent with individual abilities and preferences (a form of construct validity). Research into the predictive power of QUEST would be interesting indeed, and an evaluation of the predictive validity of QUEST would be valuable. However, introduction to occupations, not prediction, is the goal of the QUEST program. Validity of the questionnaire is therefore not contingent upon its predictive power.

Tests of concurrent validity are not required of QUEST itself, because QUEST is not a diagnostic test intended to identify latent characteristics, although certain of the items of which the Questionnaire is built are classificatory and require concurrent validity. (Again it is possible that QUEST does classify users into categories, i.e. that there are attribute and preference patterns among people which the questionnaire identifies, but that is not its essential function.) Self-reporting is presumed, for the purposes of QUEST, to be a reasonable substitute for diagnostic testing. That proposition does

require tests of concurrent validity, and the results of such tests are presented later.

The constructs upon which QUEST is based are those of traditional occupational and labor market analysis, and QUEST uses the empirical results of those analyses in relating occupations to individual attributes and preferences. For the most part, QUEST uses the worker traits identified for occupations by the U. S. Training and Employment Service and published in the third edition of the Dictionary of Occupational Titles and its supplements. Where worker traits are not used, as in the questions regarding regional and income preferences, standard labor market data are employed. The construct validity of the analytical methods and taxonomies used in the worker trait and labor market analyses upon which QUEST is based is described in the technical literature concerning those methods. Those are of sufficient conceptual credability and historical utility not to require further examination here. Additionally, by virtue of the fact that QUEST does sample a wide variety of occupationally relevant attributes, it has a kind of content validity which is lacking in even the best single-purpose test.

It was argued earlier that QUEST is a questionnaire, not a test in the usual sense of the term, and it therefore is not subject to the validation procedures prescribed for tests. Whether one accepts that distinction or not, this review of validation criteria leads to the same conclusion, namely, that the essential feature of QUEST whose validity must be demonstrated is the use of self-reporting as an operational procedure.

Self-Reporting

This questionnaire contains two types of questions. With regard to such characteristics as physical limitations, the issue is whether people know the objective fact and will accurately state whether, for instance, they can lift and carry 50-pound objects. The aptitude questions, like the physical capacity questions hinge on whether a person knows some presumably objective and potentially observable fact about himself, i.e. does he have good manual dexterity, clerical aptitude, learning ability, etc. Aptitude tests are designed to measure these inate qualities, even though the conceptually clean distinction between basic aptitudes and acquired skills presents serious measurement problems as does the heavy reliance of tests on reading skill.

The second type of question is basically an inventory of preferences, rather than immutable characteristics. With regard to working conditions, for instance, it is a question of whether people know and will report their preferences vis a vis working inside or outside, etc. On the question of educational preparation, the question again is how well people know their own minds, i.e., how much schooling are they willing to undergo by way of occupational preparation.

The interest questions are also of the second category, i.e., does the respondent know (and will he say) what kind of work activities he would or would not like to do. Some of these preference factors remain fairly stable over time and, in the case of interests, can be measured to some degree, but this fact does not change their basically preferential as opposed to innate nature.

Validation of the reporting of measurable, innate characteristics can mostly be done against some objective standard. Validation of preferential items however is much more difficult, because of their discretionary nature and the difficulty of objective measurement. At issue here is the identification of an appropriate benchmark against which to judge questionnaire responses. The best validation in most cases is probably a re-interview designed to determine whether the person understood the question and whether he stated his actual preferences. The benchmark used in this limited validation study are two: test results in the case of aptitude reporting and independent counselor assessment for other items. Limitations of these benchmarks will be discussed later.

Before examining the data concerning the concurrent validity of self-reporting, let us be clear as to the implications of our research.

One counseling psychologist (George Ralph, Psychologist, University of Oregon Counseling Center) who reviewed the Quest-Test data and this report observed that even the discrepancies which do appear between client response and independent evaluation should not be viewed as errors. Instead, he said, they should be viewed as perceptual differences on the part of the client which require re-education to bring his self perception more into line with reality. He argued that the real test of the questionnaire should be whether it accurately reflects the client as he sees himself. Counselors who have used the questionnaire say it does, and the readability analysis of Part I tends to substantiate this opinion. By helping to identify discrepancies between

the client's self perception and reality, the questionnaire becomes a highly useful counseling tool, in which the identification of discrepancies is an attribute, not a weakness. As reported following the first evaluation, several counselors who have used the questionnaire have found it very useful as such an interviewing schedule, both to raise relevant questions which might otherwise be overlooked and to create opportunities for counseling. One Employment Service counselor said of the Quest-Test:

> I found that the most useful part of the process was going back over the questionnaire with the client and discussing their answers. At times this provided a good means of getting into areas that might otherwise have not been explored -- and these were not necessarily related to the question.

This counselor is now using the questionnaire in this way, even though he does not have access to a means for translating the questionnaire responses into a list of occupations.

The questionnaire is a useful instrument, whether it elicits responses that are consistent with objective measurement or, instead, identifies perceptual errors on the part of the client. Discrepancies which occur between the applicant's responses and independent evaluation (including test results) should be viewed as perceptual error on the part of the applicant and not necessarily as a deficiency in the questionnaire. The fact that such discrepancies may be revealed by use of the questionnaire represents a useful and practical potential benefit to the counselor and to the applicant.

At issue is the appropriate operational procedure: when and for whom

are the responses close enough to reality to feed directly into QUEST processing to generate an occupational exploration list, and when should the questionnaire be used as a tool by the counselor. These are distinctly separate, though compatible, uses.

The questionnaire itself passes the test of utility whichever way it is used. A negative conclusion concerning the validity of self-reporting will not invalidate the questionnaire, only its independent use by clients. The question of validity, then, concerns the independent use of the QUEST process by clients.

Quest-Test Evaluation of Self Reporting

In the Quest-Test, counselors from the Downtown, Concentrated Employment Program (CEP), and Youth Opportunity Center (YOC) offices of the Employment Division in Portland Oregon were asked to select clients who were in the process of developing career choices and for whom information about occupations was pertinent. The client and the counselor then each completed a QUEST questionnaire about the client. (The counselor also provided some information about client characteristics: education, age, GATB scores, etc.) As reported in Part I, the client's questionnaire was then reviewed for wording problems. (A copy of the full Quest-Test administration procedures is attached.)

Quest-Test Client Characteristics

About two-thirds of the clients included in the Quest-Test were disadvantaged. Half of the Quest-Test clients had less than a high school

education. (This applied to both the disadvantaged and non-disadvantaged.) About 40 percent were youth, while most of the remainder were between 22 and 54. About 15 percent were veterans and about 15 percent were handicapped. The range of intelligence scores as measured by the GATB was completely normal, i.e., representative of the population as a whole, while the range of verbal aptitudes was biased slightly to the low end of the scale. (See Table 2 in Part I for complete data on client characteristics.)

Overall, the Quest-Test sample is rather representative of the adult population as a whole, except for being heavily weighted by disadvantaged persons.

Degree of Client-Counselor Concurrence

A "correct" response by the client is defined in the Quest-Test as one which agrees with the response estimate of the counselor or, in the case of the aptitude questions, with the General Aptitude Test Battery (GATB) scores. Thus, a client is said to be able and willing to respond to the QUEST questions if his responses agree with his counselor's or the GATB's assessment of him, and, to the extent that such agreement exists, the question is judged to show concurrent validity. (Limitations of the validation procedure are noted later.)

Overall, there was 80 percent concurrence between clients and the counselor test benchmarks. This figure and the others which follow is based on the following "scoring" procedure which not only recognizes discrepancies, but takes into account the seriousness of the discrepancy.

Scoring

After each client's questionnaire was paired with that of his counselor, a "discrepancy score" was constructed to measure client-counselor discrepancies on each of the 25 questionnaire items. (Project staff were responsible for all analysis.) A "0" was alloted when no discrepancy occurred, that is when both the client's and the counselor's answers were the same. A "1" was assigned to those responses which differed to a limited degree, e.g., a client saying he required $500 income and the counselor saying the client needed $700. A value of "2" was given to those more extreme differences which represented opposite answers or where the differences in answers amounted to 3 degrees or more. This range of discrepancy values was applicable to each question. Hence, a perfect score (i.e., complete client-counselor concurrence) would be 0, while the maximum discrepancy on each of the 25 questions would result in a score of 50.

In questions 9-13 (those pertaining to aptitudes), client responses were gauged against GATB scores whenever they were available. Although counselor assessments were available for these questions, GATB scores are presumed to be more valid measure of aptitudes.

A few additional scoring contingencies should be clarified. First of all, whenever the client or counselor selected a DK (don't know) response, in reference to some interest or attribute, a "0" discrepancy score was

ascribed to that particular question. In addition, in the interest questions either a "YES"-"NP" discrepancy was scored as "0", while a "NO"-"NP" discrepancy was scored as "1". The reason for this lies in the fact that the "NO" response represents a much stronger statement of preference in QUEST than does the "YES" response.

Table II-1 shows the degree of concurrence for each of the 25 questions. The overall average of 80 percent includes a range from a low of 62 percent (Question 8: Education) to a high of 97 percent (Question 3: Deafness). There is no discernable pattern to the differences in degree of concurrence. Of the five questions with a correspondence score below 75 percent, one is a question about working conditions (Quest 5: City Size, 70 percent concurrence); one an education question (Question 8: Education, 62 percent); one pertains to aptitudes (Question 13: Catching on to Things, 73 percent), and two are interest questions (Question 17: Interpreting facts, 74 percent; and Question 19: Copying-Sorting, 71 percent). The questions with high degrees of concurrence are similarly distributed, though the highest are both in the physical limitations section.

There is a similar lack of consistent pattern as to the type of discrepancy on most questions. For instance the discrepancies on the education question consisted about equally of over-reporting and under-reporting as to willingness to get further education.

In the self-reporting of aptitudes, however there is a distinct pattern of error. Those who misclassified themselves almost always over-estimated their aptitudes; (92 percent of the errors were over-estimates). Clients' ability and willingness to self-report their aptitudes is rather similar regardless of whether the aptitude being reported is intelligence or manual dexterity, as the data in Table II-1 indicate. Similarly, the tendency to over-report is consistent from one aptitude to another.

This pattern may not represent a biased perception so much as an unwillingness to report accurately on the part of people with low ability. Some counselors observed that Employment Service clients have an incentive to over-report abilities because of the possible use of such information in screening for training or job referral. If this institutional incentive does account for the bias toward over-reporting, it is possible that more accurate questionnaire responses would accompany completely private use of the questionnaire, where the client knew that his answers would not be seen by Employment Service staff and he would have less incentive to try to "look good." This "privacy" feature and its concomitant lack of social pressure favoring one response over another has been cited by high school and community college students as an attractive feature of the system (where use was private), and one which "permitted" them to be honest about such things. It is not known at this time, however, what pattern of reporting discrepancy is exhibited by these groups.

TABLE II-1

QUEST-TEST DEGREE OF CONCURRENCE
FOR INDIVIDUAL QUESTIONS

QUESTION	TOTAL	DISADVANTAGED	NON-DISADVANTAGED
TOTAL			
Number of clients	61	42	19
Concurrence	80 %	80 %	80 %
QUESTIONS:			
1. Lifting	82	85	76
2. Impaired Vision	89	86	95
3. Deafness	97	95	100
4. Region	75	72	82
5. City Size	70	68	74
6. Working Conditions	85	87	82
7. Sex	79	79	76
8. Education	62	61	66
9. Working with Hands	76	76	74
10. Eye for Accuracy	85	88	76
11. Ability with Words	80	75	89
12. Working with Numbers	82	81	82
13. Catching on to Things	73	74	68
14. Precision Work	82	80	87
15. Operating Equipment	83	80	92
16. Handling Materials	77	74	81
17. Interpreting Facts	74	74	74
18. Organizing, Using Information	81	84	74
19. Copying, Sorting	71	70	74
20. Counseling	86	88	82
21. Discussing, Negotiating	78	75	84
22. Instructing, Supervising	79	84	68
23. Selling, Persuading	79	82	72
24. Assisting	84	84	84
25. Monthly Wage	83	82	87

An examination of the concurrence scores for disadvantaged and non-disadvantaged shows, first, that disadvantaged clients are as able and willing as non-disadvantaged clients to answer the questions. There are, of course, differences among questions, but there are very few striking differences between the two groups. The greatest difference was a greater ability of disadvantaged clients to correctly answer Question 22 about a preference for instructing-supervising. Here disadvantaged clients concurred with their counselors' estimates 84 percent of the time; whereas non-disadvantaged clients concurred only 68 percent of the time. Other substantial differences were found in two aptitude questions: Question 11 concerning verbal ability where the non-disadvantaged did better, and Question 10 concerning clerical aptitude where the disadvantaged did better.

Besides identifying questions where clients may have trouble self-reporting, this evaluation permits us to determine which types of clients have the greatest trouble completing the questionnaire. Table II-2 shows the degree of concurrence for various client groups.

The most pronounced feature of these data is the similarity among clients of various ages, educational levels and disadvantaged status. None of these differences was statistically significant (at the .05 level). School drop-outs and disadvantaged persons apparently are as able and willing to answer the questionnaire as other people. (The characteristics of the disadvantaged in this study suggest that, while they are disadvantaged, they are not the most severely disadvantaged educationally.)

TABLE II-2:

QUEST TEST
DEGREE OF CONCURRENCE FOR CLIENT GROUPS

CLIENT CHARACTERISTICS	NUMBER OF CLIENTS IN QUEST TEST	DEGREE OF CONCURRENCE
TOTAL	61	80
Highest School Grade:		
0-7	1	82
8-11	30	76
12	22	84
Over 12	8	84
Age:		
Under 22	30	80
22-54	29	80
55+	2	78
Disadvantaged:		
YES	42	80
NO	19	80
Veteran:		
YES	9	76
NO	52	80
Handicapped:		
YES	8	76
NO	53	80
Aptitude Test Scores (GATB):		
Intelligence (G)		
Level 1	2	96
Level 2	11	80
Level 3	12	80
Level 4	11	74
Level 5	4	76

TABLE II-2 (Continued)

CLIENT CHARACTERISTICS	NUMBER OF CLIENTS IN QUEST TEST	DEGREE OF CONCURRENCE
Verbal (V)		
Level 1	3	90
Level 2	8	80
Level 3	13	80
Level 4	11	76
Level 5	5	72
Numerical (N)		
Level 1	1	90
Level 2	5	84
Level 3	14	80
Level 4	17	78
Level 5	3	70
Clerical (Q)		
Level 1	2	86
Level 2	9	82
Level 3	21	78
Level 4	7	76
Level 5	1	66
Manual (M)		
Level 1	4	86
Level 2	6	80
Level 3	10	74
Level 4	13	80
Level 5	7	76

There are some significant differences when one examines the aptitude and reading levels of clients. People with low mental aptitude levels (intelligence (G), verbal (V), numerical (N), and clerical (Q)) show consistently lower degrees of concurrence than people with average or high mental aptitudes. Physical aptitudes, at least manual dexterity, seem not to be related to ability with the questionnaire.

As can be seen from Table II-2, there is a difference of about 20 percentage points between the top tenth of the population (Aptitude Level 1) and the bottom tenth (Aptitude Level 5). The former concur with their counselors' evaluations about 90 percent of the time while the latter agree about 70 percent of the time.

More pertinent, however, is the capacity of people of low ability to handle the questionnaire independently, i.e., whether their concurrence is appreciably less than, say, people of average ability. It appears from Table II-2 that people at Level 5 (the bottom 10 percent of the population) may have more trouble than those at Level 4, but the differences are not statistically significant.

Thus, it seems appropriate to combine these two groups and to compare the combined group (the bottom third of the population) with people of average ability (Group 3: the middle third of the population.) This comparison shows the following results.

Degree of Concurrence:	G	V	N	Q
Group 3 (Middle third of the population)	80	80	80	78
Groups 4 & 5 (Bottom third of the population)	74	74	76	74
Difference (group 3-Groups 4 & 5)	-6	-6	-4	-4
Significant at .05 level	yes	yes	no	no

There is probably a significantly poorer concurrence among people with low intelligence and verbal aptitude, but the difference is only 6 percentage points, which is the equivalent of a difference on one question.

There appears also to be slightly positive correlation between concurrence on the Quest-Test and level or reading ability as measured by the Wide Range Achievement Test (WRAT). Data for 15 Quest-Test clients shows a correlation coefficient of +0.08. This sample is small and only four of those clients had WRAT scores below 8th grade, but it is noteworthy that all of the clients for whom WRAT reading scores are available were disadvantaged clients from the CEP office, and they have an average degree of concurrence. The four clients with WRAT scores below 8th grade have an average degree of concurrence of 84 percent, compared to an average of 78 percent for those with higher reading ability. It appears from these data that poor readers do not have special problems, but this sample is obviously an inadequate basis for substantial conclusions, and further research would

be required to substantiate a pattern.[18]

A fundamental question raised by all of these findings is, What is an acceptable level of discrepancy? Is the 74 percent concurrence evidenced by low ability clients high enough, or should they use the questionnaire only with the assistance of a counselor? For that matter, is the 90 percent concurrence of the high ability clients acceptable?

First, let us translate these percentages back into absolute terms. A concurrence of 74 percent means that there was discrepancy between the counselor and client 26 percent of the time. Since each questionnaire had a total value of 50 points, a discrepancy of 26 percent is the equivalent of discrepancy score of 13 points. Such a score can result from a major disagreement (e.g. opposite answers) by counselor and client on 6 of the 25 questions. It can also result from differences of minor degree (scored 1 error point) on 13 questions. Similarly, concurrence of 90 percent results from a difference of degree on five questions, or disagreement on two.

Before drawing conclusions about such data, one needs to know how good the benchmark is; that is, how much of the discrepancy can be expected to be actual "error" on the part of the client and how much of the apparent discrepancy is due to error on the part of the counselor or the test against which the clients' responses have been judged.

---

[18]The availability of reading test scores became known only after the Quest-Test procedures had been finalized and the counselors had been advised not to do additional testing solely for the purposes of the Quest-Test.

Obviously, neither counselors nor tests are infallible.

One of the comments made by counselors in reviewing their own Quest-Tests was that they were not always very sure which answer to mark for a client. (Quest Test instructions told them to complete the questionnaire on the client before discussing it with him and to put a question mark beside questions which they did not know which way to answer.) Even on questions which they did answer, several counselors said they thought the clients' responses were probably as good or better than their own, especially on such preference questions as where the client wants to live and work.

Obviously, any such error on the counselor's part would raise the stated discrepancy above the actual error. In order to assess the impact of this phenomenon, which was more a feature of the Quest-Test procedure than the questionnaire itself, several counselors were asked to review, item-by-item, the questionnaires for clients with whom they had subsequently discussed the individual answers. This review invariably led to a revised estimate by the counselor showing only about half as much discrepancy as the counselor's original estimate had shown, a substantial revision.

Some additional light can be shed on the counselor's assessment abilities by comparing the counselors estimates of client aptitudes with the client's GATB scores. Overall, counselors estimated the range of client aptitudes correctly 49 percent of the time, with a degree of concurrence of 73 percent. Clients estimates of their own aptitudes matched GATB scores 52 percent of the time

and their degree of concurrence was 75 percent.[19] Thus it appears that counselors err as frequently as clients in estimating client aptitudes.

This finding and the comments of the counselors concerning clients' ability to assess their own preferences put the concurrence factors in a new perspective. It appears now that the actual error is probably about half the discrepancy score reported in the earlier analysis of this section. Thus, it is likely that the average client answers the question correctly about 90 percent of the time, while the low ability clients, if unassisted, answer correctly about 87 percent of the time.

Other Validity Tests of Self-Reporting

There has been other research into the validity of self-reporting. A recent review of the literature concerning the validity of various self-reporting devices such as questionnaire, interviews, and personal data blanks found mixed conclusions about all three. Some studies found them to be valid instruments for gathering independently verifiable information, others did not.[20]

One of the striking features of that review is the very small number of such validation studies which have been published. The following tabulation shows the validation studies reviewed by Walsh and his analysis of their conclusions. Identified parenthetically are studies wherein the information being

---

[19]Whereas clients consistently over-estimated their aptitude, counselors under-estimated client aptitudes 58 percent of the time.

[20]W. Bruce Walsh "Validity of Self-Report" Journal of Counseling Psychology, XIV, No. 1, (January 1967), p. 18.

	Number of Studies	Number Indicating Validity Total, (LM)	Number Indicating Low Validity Total, (LM)	Ambiguous Total, (LM)
Interview	27	13 (6)	9 (5)	5 (4)
Questionnaire	7	3 (2)	4 (0)	
Personal Data Blank	5	3 (2)	2 (2)	

evaluated had something to do with labor market, personnel, or related items as opposed to political opinion, sexual behavior, and assorted other topics of self-report. Walsh's analysis shows a small majority of the studies drawing the conclusion that self-reporting is valid.[21] Roughly the same pattern appears for labor market related studies.

There is a related question whether a questionnaire is an efficacious self-report device, or whether a personal interview would be preferable, as generally assumed.

Walsh's conclusion from his review of the literature as well as his own experimentation with alternative data collection methods indicates that interviews are not more accurate than questionnaires in obtaining discrete items of information.

> No one method elicits more accurate self-report than another. In general, subjects gave quite accurate responses to most of the informational items....
>
> The results lend equal credence to research results based on the questionnaire, interview, and personal data blank.[22]

---

[21] Ibid., p. 22

[22] Ibid., p. 22

Walsh's later research substantiated this conclusion and, further, showed that subjects did not distort their answers even when given a social incentive to do so.[23]

One of the common ways of evaluating self-assessment is through comparison with test results, and one of the characteristic conclusions regarding self-assessment is that it correlates at least moderately well with test results. One study of male, student counselees found a statistically significant relationship (at the .01 level) between expressed vocational interests and interests as measured by the Strong Vocational Interest Blank.[24] An earlier study of college freshmen also found a positive relationship between estimated abilities, knowledges, and interests and the results of tests, despite wide variation among individuals and among subjects.[25]

Analyzing those individual differences further, Arsenian found that students who grossly over-estimated or under-estimated their abilities, knowledges, and adjustment tended to be somewhat less intelligent and less well adjusted.[26] Other research on occupational preference statements indicates that persons (in this

---

[23]W. Bruce Walsh, "Self-Report under Socially Undesirable and Distortion Conditions," Journal of Counseling Psychology, XVI, No. 6, pp. 569-574.

[24]Harriet A. Ross and Charles F. Elton, "Ask Him or Test Him?" Vocational Guidance Quarterly (September 1967)

[25]Seth Arsenian, "Own Estimate and Objective Measurement," Journal of Educational Psychology XXXIII (1942) pp. 291-302

[26]Ibid., p. 302

case college students) with high self-esteem are more discriminating in their choice of occupations, i.e. select occupations which will satisfy their needs. Those with low self-esteem are more accepting of non-need-satisfying occupations and seek less personal satisfaction from their occupation.[27]

This, of course, says nothing about their comparative self-assessment abilities; it implies only that low self-esteem individuals will be less particular, less demanding in response to occupational selection question.

Other studies of the correlation of test and self-report yielded similar results.

> The results presented here are in general agreement with the results obtained by other investigators and the correlation between measured and self-estimated interests approximates .50.
>
> .....
>
> The men studied here found it relatively difficult to estimate their own scientific interests, as they are measured by tests, and on the other hand, the men were more able to estimate their persuasive and sales interests.[28]

The importance of these differences in people's capacity to estimate their own interests and abilities is stressed in another article on the subject, which indicates that clients are better able to estimate those interests which are vocationally relevant. This is important for studies which compare self-reported and measured profiles. Higher degrees of accuracy may occur on test

---

[27]Abraham K. Korman, "Self-Esteem Variable in Vocational Choice." Journal of Applied Psychology, vol. 50, No. 6, p. 485

[28]Ralph F. Berdie, "Scores on the Strong Vocational Interest Blank and the Kuder Preference Record in Relation to Self Rating," Journal of Applied Psychology, XXIV, No. 1, (February 1950), pp. 42-49.

scales of high relevance, but that accuracy may be masked by poorer agreement on the more numerous but less relevant scales.[29]

Thus the significant and positive relationship between self reporting and test results might be even stronger than many studies indicate. The relevance of this factor for the OIAS questionnaire lies in the fact that every choice question has a "don't know" or "don't care" response category. Items in which respondents make affirmative responses are probably those of vocational relevance and should therefore be expected to show relatively high correlations with test results.

Counselors are always cautious about accepting verbalized statements of interest. However, this review of the literature shows that, far from repudiating self-reporting as an assessment method, recent psychological research has revealed a significant correlation between self-reporting and test-measured interests, abilities, and other personal factors. In fact, there is a growing opinion that self reporting is more than a fair proxy for test results, that it is actually a preferable assessment method in many circumstances. However it should be remembered that significant concurrence with test results is not ipso facto proof of the validity of self-reporting, since tests themselves are only approximate measures and far from infallible.

A substantial amount of research has been done on the comparative predictive validity of self-report and tests, especially interest inventories.

---

[29]Milton E. Foreman and Leonard E. James, "Vocational Relevance and Estimated and Measured Test Scores," Journal of Counseling Psychology, XVI, No. 6, pp. 547-550.

While the Occupational Information Access System questionnaire is not a predictive device, this research is pertinent because it provides more than a correlation analysis between self-reporting and test results, it provides an evaluation of both methods against an independent benchmark--subsequent behavior.[30]

While it is widely assumed that test results are preferable, research does not confirm this predisposition.

The predictive power of tests and school grades was seriously challenged in 1934 by Edward L. Thorndike. In his study Prediction of Vocational Success, he found that measured aptitudes, school grades, and school conduct (school progress, conduct, scholarship, and attendance) proved to be very poorly correlated (from zero to about .25) with early job success as measured by job level, income, or satisfaction.

> On the whole, the vocational histories of these boys and girls are not in accord with the opinions of those enthusiasts for vocational guidance who assume that an examination of a boy or girl of fourteen and a study of his school record will enable a counselor to estimate his fitness to succeed in this, that, or the other sort of work.[31]

---

[30]Once again, a critical review of the research leaves some doubt about its relevance to the issue of self-reporting by disadvantaged adults. Validation studies are rather few in number and generally use students rather than adults as subjects.

[31]Edward L. Thorndike, Prediction of Vocational Success (New York: The Commonwealth Fund, 1934), p. 72.

In a later follow-up study of the educational and occupational histories of 10,000 Air Force Aviation Cadet applicants, Thorndike and Hagen concluded:

> As far as we were able to determine from our data, there is no convincing evidence that aptitude tests or biographical information of the type that was available to us can predict degree of success within an occupation insofar as this is represented in the criterion measures that we were able to obtain. This would suggest that we should view the long-range prediction of occupational success by aptitude tests with a good deal of skepticism and take a very restrained view as to how much can be accomplished in this direction.[32]

Even the development of GATB norms for specific occupations are based on the success of persons already in the occupation, so they do not indicate the predictive power of the GATB when administered to school children or new labor force entrants.

In reviewing the entire literature on the predictive power of the Strong Vocational Interest Blank (SVIB), Dolliver concluded, "There is no evidence to show that the SVIB is superior to expressed interests. Counselors and others have apparently supposed a body of experimental evidence which does not exist."[33]

After a recent, exhaustive analysis of the data from a large number of test validity studies, Edwin Ghiselli concluded:

> while there is some evidence that tests, particularly those of intellectual and of spatial and mechanical abilities, can predict general occupational trainability, tests are far less useful in the prediction of general occupational proficiency. ...
> .....

---

[32]Robert L. Thorndike and Elizabeth Hagen, Ten Thousand Careers, (New York: John Wiley & Sons, Inc., 1959), p. 50.

[33]Robert H. Dolliver, "Strong Vocational Interest Blank Versus Expressed Vocational Interests: A Review," Psychological Bulletin, LXXII, No. 2 (1969), p. 104.

> .....Taking all jobs as a whole, then, it can be said that by and large the maximal power of tests to predict success in training is of the order of .50 and to predict success on the job itself is of the order of .35.
>
> .....
>
> .....The highest validities found in any of the single studies reviewed for this summary, studies in which more than one hundred workers were used, were .77 for training criteria and .66 for proficiency criteria. ...
>
> .....
>
> One can choose among these three descriptions of the maximal predictive power of tests, depending upon one's degree of enthusiasm for tests. Indeed, with tests and criteria of perfect reliability, it is theoretically possible to develop instruments which precisely measure the traits and abilities required by a particular job and so have perfect validity. But in view of human frailty, neither test nor criterion scores can ever be expected to have perfect reliability, nor can tests of pure abilities and traits be developed. Practically speaking, therefore, the maximal validity of tests must be taken as some lesser value. ...
>
> .....
>
> It is apparent that even the most optimistic supporter of tests cannot claim that they predict occupational success with what might be termed a high degree of accuracy. Nevertheless, in most situations tests can have a sufficiently high degree of predictive power to be of considerable practical value in the selection of personnel.[34]

Analyses of the predictive power of expressed choice have by and large generated more enthusiastic conclusions. A 19 year follow-up study of Stanford University Freshmen showed a correlation of approximately .69 between freshman choice and occupation engaged in 19 years later. That is, the validity of freshman choice is .69. Moreover, students who made an early firm choice tended to select and obtain relatively high prestige employment, whereas those who changed their choice had final occupations of appreciably lower prestige value than their original choice.[35]

---

[34]Edwin E. Ghiselli, The Validity of Occupational Atitude Tests, (New York, London, Sydney: John Wiley & Sons, Inc., 1966), pp. 124-127.

[35]Edward K. Strong, Jr., "Validity of Occupational Choice," Educational and Psychological Measurement, XIII (1953), p. 120.

Why do scientifically designed and administered tests, do so poorly when compared with expressed choice? In a review of the literature on the predictive value of expressed choice, Douglas Whitney observes,

> A person's development is, of course, a product of a process involving numerous influences and interactions.
> .....
> ...It seems plausible that when a client expresses a preference for an occupation, he does so with some awareness of aptitudes required. To whatever degree this awareness is present, which would again seem to be related to age and variety of experience, an expressed choice encompasses different information than an inventory score. In fact, this possibility of representing a number of diverse factors may explain the predictive value of expressed choice--a value which is clearly distinct from that of inventories.[36]

The QUEST questionnaire has proven to be a useful device in a variety of operational modes. The question of whether independent usage is appropriate depends, in part, on clients ability and willingness to self-report their preferences and abilities. Many Employment Service clients seem to be quite capable of doing so; the possible exceptions are those of very low ability (low GATB scores) who tend to over-report their abilities. This pattern of error was anticipated. This is further rationnale for obtaining objective test data whenever possible, especially with "disadvantaged" applicants, when an assessment of aptitudes is appropriate. This characteristic over-estimation and the resulting discrepancy between self-appraisal and objective measurement is well-known to counselors and would not in itself necessarily justify exclusion of the type of "aptitude" questions used in the OIAS questionnaire because the OIAS questionnaire is compatible in form with GATB results. In the absence of

---

[36]Douglas R. Whitney, "Predicting from Expressed Vocational Choice: A Review," Personnel and Guidance Journal, XLVIII, No. 4 (December 1969), pp. 283, 285.

aptitude test results, work experience, or significant academic achievement, the applicant's own subjective perception of himself is the only alternative available. Such self-assessments, however unreliable, may be used effectively as a starting point for discussion and exploration which will, in most instances, lead to more valid and reliable assessments, with or without formalized testing. It is far better to begin, if necessary, with subjective self-judgments, however unreliable, and attempt to proceed from there to more valid and reliable assessments which may be achieved through testing or other means, than to insist on starting with testing. That, in reality, means coercing some individuals into submitting to testing as a condition for receiving services for which, in actuality, testing may not be initially necessary or even desirable.

Other recent research on the comparative validity of tested and self-reported abilities also support the use of self-reporting as at least one of the bases for questionnaire response.